AD-A097 613 CLEMSON UNIV SC DEPT OF MATHEMATICAL SCIENCES F/G 12/1
A GOODNESS-OF-FIT TEST BASED ON SPACINGS.(U)
JUL 80 K ALAM, K M SAXENA N00014-75-C-0451
UNCLASSIFIED N119 NL

1 of 1
AD
A097613

END
DATE
FILMED
5 81
DTIC

A GOODNESS-OF-FIT TEST

BASED ON SPACINGS

Khursheed Alam
Clemson University

K. M. Lal Saxena
University of Nebraska-Lincoln

Department of Mathematical Sciences
Clemson University

Technical Report #342

~~June,~~ 1980

11 3 Jul 81

N119

This work was supported by the Office of Naval Research
under Contract N0014-75-0451

A GOODNESS-OF-FIT TEST BASED ON SPACINGS

Khursheed Alam and K. M. Lal Saxena

Clemson University and University of Nebraska-Lincoln

ABSTRACT

The difference between consecutive order statistics from a sample is called a spacing. Various tests based on sample spacings have been considered in the literature for testing the hypothesis that the sample is drawn from a specified distribution. Tests based on the spacings are recommended for use when the alternative distribution differs from the hypothetical distribution in the shape of the density function. In this paper, we consider a test based on the spacings designed for the case when the ratio of the two density functions is a piece-wise monotone function. This paper deals mainly with the large sample properties of the test.

Key words: Spacings, Goodness-of-fit, Asymptotic Relative Efficiency.

AMS Classification: 62G10, 62G30

This work was supported by the Office of Naval Research under Contract N0014-75-0451.

1. Introduction. Let $X_1, \ldots, X_n$ be a sample drawn from a certain distribution. In this paper, we consider a test of the hypothesis H_o that the sample comes from a known distribution F, say, against the alternative hypothesis H_1 that the sample is drawn from a distribution G, say, where G is not known completely. Let A denote the common support of the distributions F and G. We assume that A is a finite or infinite interval and that distributions have continuous density with respect to the Lebesgue measure. Let f and g denote the respective density functions. We further assume that the ratio $Q(x) = g(x)/f(x)$ is a piece-wise strictly monotone function of x inside A. Suppose that the slope of the graph of $Q(x)$ changes sign at k points. Let $\underline{\eta} = (\eta_1, \ldots, \eta_k)$ denote the change points. We consider two cases: (i) $\underline{\eta}$ is known and (ii) $\underline{\eta}$ is not known.

We shall consider in length the case in which there is a single change of sign (k=1) in the slope of the graph of $Q(x)$. This is realized in the following examples, where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ denotes the standard normal density function.

(a) $f(x) = \phi(x)$; $\quad g(x) = (1/\sigma)\phi(x/\sigma)$

(b) $f(x) = \phi(x)$; $\quad g(x) = e^{-x}/(1+e^{-x})^2$

(c) $f(x) = \phi(x)$; $\quad g(x) = p\phi(x-\mu_1)+(1-p)\phi(x-\mu_2)$

$$\mu_1 < 0 < \mu_2, \quad 0 < p < 1$$

(d) $f(x) = 1,\ 0<x<1$; $\quad g(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

$$0 < x < 1; \quad \alpha,\beta > 1.$$

Let $Y_i = F(X_i)$, $i = 1,\ldots,n$ and let $D_1 = Y_{(1)}$, $D_2 = Y_{(2)} - Y_{(1)}, \ldots, D_n = Y_{(n)} - Y_{(n-1)}$ denote the sample spacings of the transformed data where $Y_{(i)}$ denotes the ith smallest value among $Y_1,\ldots,Y_n$. Our test is based on the spacings. The test statistic will be denoted by T. Various tests for goodness-of-fit, based on spacings have been considered in the literature. The papers of Pyke (1965, 1972), Proschan and Pyke (1967), Sethuraman and Rao (1970), Kale (1969) and Kirmani and Alam (1974) may be cited for reference. Pyke (1972) points out that tests based on spacings should be used when the alternative distribution differs from the hpothetical distribution in the shape of the density function.

The following statistics have been proposed in the literature for a test of goodness-of-fit:

$$U = \sum_{i=1}^{n} D_i^r; \quad r \quad -\tfrac{1}{2},$$

$$V = \sum_{i=1}^{n} |nD_i - 1| \quad \text{and}$$

$$W = \sum_{i=1}^{n} \log D_i .$$

The null hypothesis is rejected when the absolute value of the statistic is large. It is known (see e.g., Cibisov (1961)) that the asymptotic efficiency of any test symmetric in the spacings is equal to zero relative to the Kolmogorov-Smirnov test. Sethuraman and Rao (1970) have compared the relative efficiencies of the tests based on U, V and W.

For applications to reliability and life-testing, Proschan and Pyke (1967) have considered a test of the hypothesis that the given sample comes from an exponential distribution which has a constant failure rate property against the alternative hypothesis that the distribution has monotone failure rate. The test is based on the statistic

$$(1.1) \qquad S = \sum_{i=1}^{n} \sum_{j=i}^{n} h(\bar{D}_i, \bar{D}_j)$$

where h is a bounded nonnegative function and $\bar{D}_i = (n-i+1)D_i$ denotes the ith normalized spacing. The authors have shown that the distribution of S is asymptotically normal under the alternative hypothesis. Further, Bickel and Doksum (1969) have shown that the asymptotic normality holds also for a sequence of alternatives G_n approaching the exponential distribution.

The test statistic T is derived from S as follows: First suppose that $\underline{\eta}$ is known. Let $\xi_i = F(\eta_i)$ and let π_i denote the set of values of Y_j for which $\xi_{i-1} \leq Y_j < \xi_i$, and n_i denote the number of elements in π_i, $i = 1,\ldots,k+1$, where $\xi_o = 0$ and $\xi_{k+1} = 1$. Let Y_i^* denote the smallest value in π_i, and let $S_m = Y_i^* - \xi_{i-1}$ for $m = n_1+\ldots+n_{i-1}+1$ $(i = 1,\ldots,k+1)$ and $S_m = D_m$ otherwise, $m = 1,\ldots,n$, where $n_o = 0$. Let

$$(1.2) \qquad h(x,y) = \begin{cases} 1 & \text{for } x \leq y \\ 0 & \text{for } x > y. \end{cases}$$

Define

$$T_o = \sum_{j=1}^{n} \sum_{i=1}^{j} h(S_i, S_j), \quad \overline{T}_o = \sum_{j=1}^{n} \sum_{i=j}^{n} h(S_i, S_j)$$

$$T_i = \sum_{j=n_1+\ldots+n_{i-1}+1}^{n_1+\ldots+n_i} \; \sum_{i=n_1+\ldots+n_{i-1}+1}^{j} h(S_i, S_j)$$

$$\overline{T}_i = \sum_{j=n_1+\ldots+n_{i-1}+1}^{n_1+\ldots+n_i} \; \sum_{i=j}^{n_1+\ldots+n_i} h(S_i, S_j).$$

Let $\delta_j = 0(1)$ for odd (even) values of $j = 1,\ldots,k+1$. We let

$$T = \sum_{j=1}^{k+1} ((1-\delta_j)T_j + \delta_j \overline{T}_j) \quad \text{or} \quad \sum_{j=1}^{k+1} (\delta_j T_j + (1-\delta_j)\overline{T}_j)$$

according as the graph of $Q(x)$ has initially a positive or negative slope. If the slope does not change sign, we let

$$T = T_o \text{ or } \overline{T}_o$$

according as the slope is positive or negative. The hypothesis H_o is rejected for small values of T.

In case (ii), we estimate $\underline{\eta}$ from the data and substitute the estimate for $\underline{\eta}$ in the definition of T, given above. The estimation of $\underline{\eta}$ is considered in Section 5 below.

In Section 2, we show that the given test is unbiased. To compute the critical value of T and the power of the test, we need to find the distribution of T under H_o and H_1. The distribution is shown in Sections 3 and 4. Some results on the relative efficiency of the test are given in Section 6.

2. <u>Unbiasedness of the test</u>. Let $H = GF^{-1}$, $Y_i' = H(Y_i)$ and $D_i' = Y'_{(i)} - Y'_{(i-1)}$, $i = 1,\ldots,n$, where $Y_o' = 0$. Suppose that the slope of the graph of $Q(x)$ is positive inside (η_{m-1}, η_m). Then H is convex on (ξ_{m-1}, ξ_m). Therefore, by the mean value theorem we have that for $n_1+\ldots+n_{m-1}+1 \le i \le j \le n_1+\ldots+n_m$

$$\frac{D_j'}{D_i'} = \frac{Y'_{(j)}-Y'_{(j-1)}}{Y'_{(i)}-Y'_{(i-1)}} = \frac{Y_{(j)}-Y_{(j-1)}}{Y_{(i)}-Y_{(i-1)}} \cdot \frac{H'(C_j)}{H'(C_i)}$$

$$\text{(2.1)} \qquad \ge \frac{Y_{(j)}-Y_{(j-1)}}{Y_{(i)}-Y_{(i-1)}} = \frac{D_j}{D_i}$$

where $Y_{(j-1)} < C_j < Y_{(j)}$ and $Y_{(i-1)} < C_i < Y_{(i)}$. The inequality in (2.1) follows from the fact that $H'(C_j)/H'(C_i) \ge 1$ due to the convexity of H on (ξ_{m-1}, ξ_m). From (2.1) we find that $D_i \le D_j \Rightarrow D_i' \le D_j'$. Therefore $h(D_i, D_j) \le h(D_i', D_j')$. Let T_m', $\bar{T}_m'$ and T' be obtained from T_m, $\bar{T}_m$ and T respectively by substituting Y_i' for Y_i, $i = 1,\ldots,n$. Then by the above inequality we have that $T_m \le T_m'$. Similarly, if the slope of the graph of $Q(x)$ is negative inside (η_{m-1}, η_m) then H is concave on (ξ_{m-1}, ξ_m) and therefore $\bar{T}_m \le \bar{T}_m'$. It follows that $T \le T'$. Hence

$$P(T \le t|G) \le P(T' \le t|G) = P(T \le t|F), \quad \text{for all } t \ge 0.$$

Since H_o is rejected for small values of T, the test is unbiased.

3. <u>Distribution of T under H_o</u>. When $G = F$, the sample spacings $D_1,\ldots,D_n$ are jointly symmetrically distributed according to the Dirichlet distribution given by the density function

(3.1) $$p(d_1,\ldots,d_n) = n! \;;\quad d_i \geq 0,\; 0 \leq d_1+\ldots+d_n \leq 1.$$

Let $Z_1,\ldots,Z_n$ be n random variables jointly symmetrically distributed and let

$$R_j = \sum_{i=1}^{j} h(Z_i,Z_j), \qquad j = 1,\ldots,n$$

denote the left sequential ranks where the function h is defined as in (1.2). Similarly, let

$$\bar{R}_j = \sum_{i=j}^{n} h(Z_i,Z_j), \qquad j = 1,\ldots,n$$

denote the right sequential ranks. It is known (see, e.g., Renyi (1962)) that $R_1,\ldots,R_n$ are statistically independent and that the distribution of R_j is uniform, given by

(3.2) $$P(R_j = m) = \frac{1}{j}\,, \quad m = 1,\ldots,j.$$

The same result holds for $\bar{R}_1,\ldots,\bar{R}_n$.

The above result is easily generalized as follows: Let $(Z_1,\ldots,Z_{\ell_1}),\ (Z_{\ell_1+1},\ldots,Z_{\ell_1+\ell_2}),\ldots,(Z_{\ell_1+\ldots+\ell_k+1},\ldots,Z_{\ell_1+\ldots+\ell_{k+1}})$ be a partition of $\underline{Z} = (Z_1,\ldots,Z_n)$ into k+1 sub-vectors, where $\ell_1+\ldots+\ell_{k+1} = n$ and let

$$R_j^m = \sum_{i=1}^{j} h(Z_{\ell_1+\ldots+\ell_{m-1}+i}, Z_{\ell_1+\ldots+\ell_{m-1}+j})$$

be the jth left sequential rank of the variables in the mth partition $(Z_{\ell_1+\ldots+\ell_{m-1}+1},\ldots Z_{\ell_1+\ldots+\ell_m})$; $j = 1,\ldots,\ell_m$;

$m = 1,\dots,k+1$, where $\xi_o = 0$. Let the right sequential ranks $\bar{R}_j^m$ be defined in analogous manner. Then the random variables R_j^m, $j = 1,\dots,\ell_m$, $m = 1,\dots,k+1$ are statistically independent and the distribution of each R_j^m is uniform. Similarly, the random variables $\bar{R}_j^m$, $j = 1,\dots,\ell_m$; $m = 1,\dots,k+1$ are statistically independent and the distribution of each $\bar{R}_j^m$ is uniform. Moreover, R_i^m and $\bar{R}_j^{m'}$ are independent for $m \neq m'$. These results follow from the property of symmetry of the joint distribution of $Z_1,\dots,Z_n$.

Given $\underline{\xi}$ and $\underline{n} = (n_1,\dots,n_{k+1})$, π_i represents a sample of n_i observations from a uniform distribution on (ξ_{i-1},ξ_i). Moreover, the k+1 subsamples are conditionally independent. The following properties of the conditional distribution of $T_1,\dots,T_{k+1}$; $\bar{T}_1,\dots,\bar{T}_{k+1}$, given $\underline{n}$ follow from the results given above: (a) $T_1,\dots,T_{k+1}$ are independent, (b) $\bar{T}_1,\dots,\bar{T}_{k+1}$ are independent, (c) T_i and $\bar{T}_j$ are independent for $i \neq j$ and (d) T_i and $\bar{T}_i$ have the same distribution for each i. Therefore conditionally,

$$T \overset{d}{=} \sum_{i=1}^{k+1} T_i \tag{3.3}$$

where $\overset{d}{=}$ means "distributed as".

Let $P_{i,n_i}(t) = P(T_i = t|\underline{n})$, $i = 1,\dots,k+1$. Clearly, $P_{i,n_i}(t) = 0$ for $t < n_i$. For $t \geq n_i$, the probability can be computed recursively from the relation

$$n_i P_{i,n_i}(t) = P_{i,n_{i-1}}(t-1)+P_{i,n_{i-1}}(t-2)+\dots+P_{i,n_{i-1}}(t-n_i). \tag{3.4}$$

Since $T_1, \ldots, T_{k+1}$ are conditionally independent given $\underline{n}$, from (3.3) we get

$$(3.5) \qquad P(T = T|\underline{n}) = \sum_{t_1+\ldots+t_{k+1}} \prod_{i=1}^{k+1} p_{i,n_i}(t_i).$$

As $\underline{n}$ is distributed according to the multinomial distribution with the associated probability vector $\underline{p} = (p_1, \ldots, p_{k+1})$, $p_i = \xi_i - \xi_{i-1}$, $i = 1, \ldots, k+1$, we obtain

$$P(T = t) = \sum_{n_1+\ldots+n_{k+1} = n} P(T = t|\underline{n}) \; n! \prod_{i=1}^{k+1} \frac{p_i^{n_i}}{n_i!} .$$

Kendall (1938) and Mann (1945) have tabulated the distribution of $p_{i,n_i}(t)$, for $n_i \leq 10$.

Since T_i is the sum of independent random variables $R_1^i, \ldots, R_{n_i}^i$, it follows from Liaponouv's theorem that for large values of n_i, T_i is asymptotically normally distributed with mean $n_i(n_i+3)/4$ and variance $n_i(n_i-1)(2n_i+5)/72$. Therefore, T is asymptotically normally distributed for large n, with mean μ_n and variance σ_n^2, given by

$$(3.6) \qquad \mu_o = \sum_{i=1}^{k+1} n_i(n_i+3)/4 \sim (n^2/4) \sum_{i=1}^{k+1} p_i^2$$

$$(3.7) \qquad \sigma_o^2 = \sum_{i=1}^{k+1} n_i(n_i-1)(2n_i+5) \sim (n^3/36) \sum_{i=1}^{k+1} p_i^3.$$

where $\sim$ means "asymptotically equivalent to".

4. <u>Distribution of T under H_1</u>. The small sample theory of the distribution of T under H_1 is mathematically intractable. Therefore, we consider the asymptotic theory. Let $Z_1, \ldots, Z_n$ be a sample from the exponential distribution whose cdf is given by $A(z) = 1-\exp(-z)$, and let $Z_{(i)}$ denote the ith order statistic in the sample. Similarly, let $U_{(i)}$ denote the ith order statistic in a sample of n observations from the uniform distribution in (0,1). It is known that

$$Z_{(i)} \overset{d}{\sim} \sum_{j=1}^{i} Z_j/(n-j+1).$$

Let

$$\lambda(u) = (1-u)/H'(H^{-1}(u)),$$

where $H = GF^{-1}$. We have

$$\text{(4.1)} \qquad nD_i \overset{d}{\sim} n(H^{-1}A(Z_{(i)}) - H^{-1}A(Z_{(i-1)}))$$

$$= n(Z_{(i)} - Z_{(i-1)})e^{-z}/g(G^{-1}(A(z))),$$

by mean value theorem

$$= n(Z_{(i)} - Z_{(i-1)})\lambda(u)$$

$$\overset{d}{\sim} (n/(n-i+1))Z_i\lambda(\theta_i)$$

where $Z_{(i-1)} \le z \le Z_{(i)}$, $u = A(z)$ and $U_{(i-1)} \le \theta_i \le U_{(i)}$. Though (Z_i, Z_j) and (θ_i, θ_j) are statistically dependent, Proschan and Pyke (1967) have shown that the degree of dependence is negligible in relation to the distribution of T for

for large n. Under the condition of independence, we have

$$Eh(D_i,D_j) = P(D_i \leq D_j)$$

$$= E\,\frac{\lambda(w_j)}{n-j+1}\left[\frac{\lambda(w_i)}{n-i+1}+\frac{\lambda(w_j)}{n-j+1}\right]^{-1}.$$

Then for large n, we have

$$ET_i = E\sum_{n_{i-1}+1\leq i\leq j\leq n_i} h(D_i,D_j)$$

$$(4.2)\qquad \approx n^2\int_{\xi_{i-1}}^{\xi_i}\int_u^{\xi_i}\frac{H'(H^{-1}(u))}{H'(H^{-1}(u))+H'(H^{-1}(v))}\,dudv$$

$$= n^2\mu_i,\ \text{say}.$$

From Theorem 4.2 of Proschan and Pyke (1967) it follows that T_i is asymptotically normally distributed. The mean of the asymptotic distribution is given by (4.2). The variance of the asymptotic distribution equal to $n^3\sigma_i^2$, say, can be obtained from formulas (4.59) and (4.60) of the paper. We do not give the expression for the variance, since it is involved. The distribution of T is asymptotically normal with mean equal to $n^2\sum_{i=1}^{k+1}\mu_i$ and variance equal to $n^3\sum_{i=1}^{k+1}\sigma_i^2$.

5. <u>Estimation of $\underline{\mu}$</u>. We have considered the case when $\underline{\mu}$ or equivalently $\underline{\xi}$ is known, giving the points of inflection in the

graph of $Q(x)$. If ξ is unknown, we estimate ξ as follows. We shall describe the method of estimation when there is a single point of inflection ξ_1, though the method carries through to the case in which there are several points of inflection. Let

$$L_m = \sum_{j=1}^{m} \sum_{i=1}^{j} h(D_i, D_j) + \sum_{j=m+1}^{n} \sum_{i=j+1}^{n} h(D_i, D_j)$$

and let m^* denote the value of m, maximizing (minimizing) L_m as m varies from 0 to 1, if the slope of $Q(x)$ changes sign from negative to positive (positive to negative). If the sign changes from negative to positive then $r(u)/(1-u)$ is first increasing then decreasing as u varies from 0 to 1. From the representation (4.1) of the sample spacings, it is seen that the values of D_i tend to increase then decrease as i varies from 1 to n and therefore the value of m^* maximizing L_m is approximately given by

$$(5.1) \qquad m^* = [\xi_1 n]$$

where $[x]$ denotes the nearest integer value of x. If the slope of $Q(x)$ changes sign from positive to negative then m^* minimizes L_m. The estimate of ξ_1 is given by the largest value of ξ_1 satisfying the relation (5.1).

It can be shown that $m^*/n = \xi_1 + 0_p(m^{-1/2})$. Therefore, the asymptotic theory developed in the previous section remains valid when ξ_1 is replaced by its estimated value in the definition of T.

6. Asymptotic relative efficiency. In this notion, we compare the test based on T with a likelihood ratio test, using the criterion of asymptotic relative efficiency (ARE) for the comparison. We consider below two examples for the comparison. In Example 1 we test an exponential distribution agains a Weibull distribution. In Example 2 we test a uniform distribution against a beta distribution. For a specified set of alternatives indexed by θ, say, the formula for the ARE of a sequence of tests (based on a sequence of asymptotically normal test statistics $\{T_n\}$) against a sequence of tests (based on the asymptotically normal test statistics $\{t_n\}$) is given by the formula (see e.g., Gibbons (1971))

$$\text{(6.1)} \qquad \text{ARE} = \lim_{n\to\infty} \left[\frac{\mu'_{T_n}(\theta_o)}{\sigma_{T_n}(\theta_o)} \Big/ \frac{\mu'_{t_n}(\theta_o)}{\sigma_{t_n}(\theta_o)} \right]^2$$

where θ_o denotes the null hypothesis, $\mu_{T_n}(\theta)$ and $\sigma^2_{T_n}(\theta)$ denote the limiting mean and variance, respectively, of $\{T_n\}$, $\mu'_{T_n}(\theta)$ denotes the derivative of $\mu_{T_n}(\theta)$ with respect to θ. The parametric functions $\mu_{t_n}(\theta)$, $\sigma^2_{t_n}(\theta)$ and $\mu'_{t_n}(\theta)$ are defined similarly, as for $\{T_n\}$.

Example 1. Let $f(x) = e^{-x}$, $x > 0$ and $g(x) = \theta x^{\theta-1} e^{-x^\theta}$, $x > 0$, $\theta > 1$. It is seen that $Q(x)$ is increasing (decreasing) in x for $x \leq (>)$ for all $\theta > 1$. We have $\xi_1 = F(1) = 1-e^{-1}$. The spacing-test statistic is

$$T = T_1 + \bar{T}_2 \ .$$

From (4.2) we have, after simplification,

$$\mu_T'(1) \approx n^2(I_1+I_2) \tag{6.2}$$

where I_1 and I_2 are given by

$$I_1 = \iint_{0<y<z<1} \frac{y^2z^2e^{-y-z}[(1+y)\log y - (1+z)\log z]}{(y^2+z^2)^2}\,dydz \tag{6.3}$$

$$= \qquad\qquad , \text{ approximately.}$$

$$I_2 = \iint_{0<y<z<1} \frac{y^2z^2e^{-(y+z)/yz}[((1+z)/z)\log z-((1+y)/y)\log y]}{(y^2+z^2)^2}\,dydz \tag{6.4}$$

$$= \qquad\qquad , \text{ approximately.}$$

The approximate values of I_1 and I_2 given above are obtained by numerical integration. From (3.7) we have

$$\sigma_T^2(1) \approx \frac{n^3}{36}(1-2\xi_1(1-\xi_1)). \tag{6.5}$$

Proschan and Pyke (1967) have shown that the likelihood ratio test rejects the null hypothesis when T_W is large where

$$T_W = \sum_{i=1}^{n}(1-X_i)\log X_i \tag{6.6}$$

and

$$\mu_{T_W}'(1) = n[(\gamma-1)^2+\pi^2/6],\ \gamma = .5772, \text{ approximately.} \tag{6.7}$$

(6.8) $$\sigma^2_{T_W}(1) = n[(\gamma-1)^2+\pi^2/6].$$

From (6.1) the ARE is given by

$$ARE = 36(I_1+I_2)^2/(1-2\delta_1(1-\delta_1))((\gamma-1)^2+\pi^2/6)$$

$$= \qquad\qquad \text{approximately.}$$

Example 2. Let $f(x) = 1$, $0 < x < 1$ and $g(x) = \theta x^{\theta-1}$, $\theta > 1$. It is seen that $Q(x)$ is increasing in x inside the interval $(0,1)$. Therefore $T = T_o$. From (4.2) and (3.7) we get after some simplification

$$\mu'_{T_o}(1) \approx -\frac{n^2}{2}\left[\frac{1}{3^2} - \frac{2}{5^2} + \frac{3}{7^2} - \ldots\right] \approx -(.031)n^2$$

$$\sigma^2_{T_o}(1) \approx n^3/36.$$

The likelihood ratio test rejects the null hypothesis when $T_B = \sum_{i=1}^{n} \log x_i$ is large. By direct computation

$$\mu'_{T_B}(1) = -n \quad\text{and}\quad \sigma^2_{T_B}(1) = n.$$

From (6.1) the value of the ARE turns out to be equal to .0345 approximately.

References

Bickel, P.J. and Doksum, K.A. (1969). Test for monotone failure rate based on normalized spacings. Ann. Math. Statist. 40, 1216-1235.

Cibisov, D.M. (1961). On the test of fit based on sample spacings. Theoria vero. i. Prim. 6, 354-8, Theory Probab. Applic. 6, 325-9.

Gibbons, Jean D. (1971). Nonparmetric Statistical Inference. McGraw Hill Book Co. New York.

Kale, B.K. (1969). Unified derivation of tests of goodness of fit based on spacings. Sankhya, Series A, 31, 43-48.

Kendall, M.G. (1938). A new measure of rank correlation. Biometrika, 30, 81-93.

Kirmani, S.N.U.A. and Alam, S.N. On goodness of fit tests based on spacings. Sankhya, Series A, 36, 197-203.

Mann, H. B. (1945). Nonparametric tests against trend. Econometrica 13, 245-249.

Proschan, F. and Pyke, R. (1967). Tests for monotone failure rate. Proc. Fifth Berkley Symp. on Math. Stat. and Probab. Univ. of California Press. 3, 293-312.

Pyke, R. (1965). Spacings. J. Roy. Statist. Soc. Ser. B. 27, 395-449.

Pyke, R. (1972). Spacings revisited. Proc. Sixth Berkley Symp. on Math. Stat. and Probab. Univ. of California Press. 1, 417-427.

Renyi, A. (1962) Theoré des elements d'une suité d'observations, Colloquium on Combinatorial Methods of Probability Theory. Aarhus Univ. 104-115.

Sethuraman, J. and Rao, J.S. (1970). Pitman efficiencies of tests based on spacings. Nonparametric Techniques in Statistical Inference, Ed. M.L. Puri. 405-416, Cambridge University Press.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: N119	2. GOVT ACCESSION NO.: AD-A097 613	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): A Goodness-of-Fit Test Based on Spacings		5. TYPE OF REPORT & PERIOD COVERED: Technical
		6. PERFORMING ORG. REPORT NUMBER: Technical Report #342
7. AUTHOR(s): Khursheed Alam and K. M. Lal Saxena		8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0451
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Clemson University, Dept. of Mathematical Sciences, Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-271
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Code 436, Arlington, Va. 22217		12. REPORT DATE: July 3, 1980
		13. NUMBER OF PAGES: 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spacings, Goodness-of-fit, Asymptotic Relative Efficiency.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The difference between consecutive order statistics from a sample is called a spacing. Various tests based on sample spacings have been considered in the literature for testing the hypothesis that the sample is drawn from a specified distribution. Tests based on the spacings are recommended for use when the alternative distribution differs from the hypothetical distribution in the shape of the density function. In this paper, we consider a test based on the spacings designed for the case when the ratio of the two density functions is a piece-wise monontone function. This paper deals mainly with the large sample properties of the test.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DATE
FILMED

8